A systematic review and meta-analysis on the impact of proficiency-based progression training on trainees' performance outcomes in medicine.

Published in:
Annals of Surgery

Publication Status:
Accepted/In press: 13/11/2020

Document Version
Author Accepted version

A systematic review and meta-analysis on the impact of proficiency-based progression training on trainees' performance outcomes in medicine.

Corresponding Author:

Words count (Abstract): 336 (Max 350)

Words count (manuscript): 3,504 (Max 3500)

Tables: 2

Figures: 3

Key words: Surgical training; proficiency-based progression training; proficiency-based metrics; objective performance metrics; procedural errors; procedural steps; technology enhanced training; simulation-based training.

ABSTRACT

Importance: To date, the impact of 'proficiency-based progression' (PBP) methodology to learning clinical skills in comparison to the traditional approach to training has not been reviewed and analyzed systematically.

Objective: To systematically analyze all published prospective, randomized and blinded clinical studies on the PBP training methodology.

Data Sources: Comprehensive search of PubMed, Cochrane library's Central, EMBASE, MEDLINE and Scopus databases, from their inception to 1st March 2020. All the references identified from bibliographies of key reviews on training were also screened.

Study Selection: Inclusion criteria were studies using objective performance metrics and PBP methodology

Data Extraction and Synthesis: Two independent~~s~~ reviewers ~~abstracted~~ extracted the data. The Medical Education Research Study Quality Instrument (MERSQI) was used to assess the methodological quality of the included studies. The risk of bias for all studies was assessed by three independent investigators and their inter-rater reliability (IRR) was calculated.

Main Outcome(s) and Measure(s): The primary outcome was the number of procedural errors performed comparing PBP and non-PBP-based training pathways. Secondary outcomes were the number of procedural steps completed and the time to complete the task/procedure. Results were pooled using biased corrected standardized mean difference (SMD) and ratio-of-means (ROM) using random-effects models. (Is the biased correction Hedges' g?) (Is the ratio of means the same as the response ratio, and if so does one not require a ratio scale?) (I don't get the link between ROM and a random effects model.) (In a random-effects model the effect size varies depending on the study,

unlike the fixed effects model where it is assumed that all of the studies share a common effectsize.)

Results: From the initial pool of 468 studies a total of~~Overall,~~ 12 studies, with a total of ~~enrolling~~ 239 participants, were included in the current study. ~~from an initial pool of 468 studies.~~ The mean MERSQI score of the included studies was high (15.5). When comparing~~ed~~ ~~to~~ standard simulation based training to~~,~~ PBP training a reduction in~~ed~~ the number of errors was reported (SMD -2.68, 95%CI: -3.52; -1.83; $p < 0.001$) and procedural time was also reduced (SMD -0.93, 95% CI: -1.55; -0.30; $p = 0.003$), while ~~and increased~~ the number of performed steps increased (SMD 3.46, 95% CI: 2.13; 4.79; $p < 0.001$). Using a ~~At~~ ROM comparison~~analyses,~~ PBP was estimated to reduce~~/increase~~ the mean number of errors by 58% ~~and steps~~ and procedural time taken by 15%, while on average ~~of, respectively, -58%, +43% and -15%~~increasing the number of steps taken by 43% when compared to standard training. As a test of sensitivity a series of s~~S~~ubgroup analyses were conducted~~in studies~~ based on ~~with~~ intraoperative performance assessments and these supported the above results. ~~confirmed all the reported findings.~~ (How do we know that a MERSQI score of 15.5 is high?) (Maybe something more needs to be said about the subgroup analyses.)

Conclusions and Relevance: Our systematic review and meta-analysis confirms that PBP training improves trainees' performances, by decreasing procedural errors and procedural time, while increasing the number of correct steps taken~~by 60%,~~ when compared to standard simulation-based training.

1. Introduction

During his time as a program manager at the Defence Advanced Research Projects Agency (or DARPA) Satava, a US Army surgeon, proposed that surgeons acquire~~d~~ their skills for procedures such as laparoscopic cholecystectomy outside of the operating room ~~and~~ on a~~of~~ virtual reality simulator.[1] Although a new concept in surgery, simulation-based training had a strong foothold in other safety conscious industries such as aviation,[2] nuclear-power,[3] and had been used in anaesthesia for more than a decade.[4]

The aim of training sessions in anaesthesia was to give the individuals or teams the experience of emergency situations before they were actually encountered in a real-life clinical situation. In contrast, Satava proposed procedural-based skills training on a virtual reality simulator~~ion~~. In 2002 the first prospective randomised and blinded clinical study of simulation-based training for the operating room demonstrated that simulation-based surgical trainees performed significantly better than traditional, trained surgeons whilst performing part of laparoscopic cholecystectomy on real patients.[5]

The methodology used in this clinical study differed significantly from previous studies. The operative procedure (i.e., dissection of gallbladder from the liver-bed with electrocautery) was characterised in detail to identify intraoperative performance metrics for optimal and suboptimal (i.e., deviations from optimal performance or 'errors') performance.[6] The metrics were explicitly defined and attending surgeons were trained to score them reliably (i.e., with an interrater reliability > 0.8).[7–9] A well validated virtual reality (VR) simulator [10–14] was then used to train medics in the technical skills required to perform a given ~~of the~~ operative procedure. Unlike previous studies, the trainees were required to continue~~d~~ training until they could demonstrate~~d~~ that they~~a~~ had met the requirements of a quantitatively pre-defined ~~performance~~ benchmark or proficiency level. This level of proficiency ~~which~~ was based on the mean performance of the attending surgeons conducting the ~~on~~ the same tasks, and on the same VR Simulator.[15]

During the next two decades the 'proficiency-based progression' (PBP) methodology evolved in terms of the robustness of the metric development and validation.[16,17] The metrics were derived from experienced and practising clinicians and ~~; they~~ represented the baseline~~a~~ reference criteriaapproach ~~(i.e., uncomplicated and straightforward) to~~ for a given~~the~~ procedure. P~~p~~erformance criteria fo~~r a~~ trainees~~ at the start of their learning curve rather than every procedure conceivable; they~~ were developed by a small group and then presented to a Delphi panel of peers for review and consensus[16,18,19] followed by further tests of quantitative validation~~ efforts such as construct validity~~.[18,20–23] ~~Thus, the simulations were derived from and based upon the metrics rather than the other way around.~~ ~~Furthermore, w~~Where~~n~~ a VR simulator~~ion~~ was not available the metrics were developed ~~used in conjunction with available~~ using simulation models; e.g., knot tying models,[24,25] silicon models[26] or cadavers.[23] A simulation can be defined as an artificially created or configured ‘learning’ situation that allows for the practice or rehearsal of all or salient aspects of a procedure. Any such ~~The~~ artificial learning situation should provide the span of appropriate sensory responses to the learner and be consistent with physical actions that are behaviourally consistent with what would be experienced in real life (including the opportunity to enact both appropriate and inappropriate learner actions (i.e., errors)). The simulation should also afford the opportunity to perform the procedures in the same order, and with the same devices, as in the real situations~~ that the procedure would normally be performed~~. Crucially, the simulation methodology ~~it~~ should ~~also~~ afford reliable and valid metric-based assessment of performance. Assessments must, at a minimum, allow summative, but preferably formative feedback, on the performance of the procedure~~ performance~~ proximate to task execution, particularly for metric errors.[27,28] This configuration allows for the trainee to engage in deliberate rather than repeated practice which is a more effective way to learn skilled performance.[29]

The requirement to demonstrate to a ~~the~~ quantitatively pre-defined proficiency benchmark ~~before progression~~ in training, combined with ~~deliberate practice~~ simulation based practice, meant that~~e~~ PBP training was particularly effective; ~~and~~ demonstrated performance improvements >40%

in objectively assessed intra-operative errors in comparison to traditional skills based training in the areas of laparoscopic surgery,[5,15,30] arthroscopic surgery[31], endovascular interventions,[32] anaesthesia,[33] and communication skills for deteriorating patients.[34]

Several focused reviews have attempted to delineate the impact of simulation-based training specifically for laparoscopic surgery[35,36]. H~~h~~owever, each had limitations including ambiguous classification of comparison interventions, incomplete assessment of study quality, or no quantitative pooling to derive best estimates of effect or effect size. A more recent review concentrated on the impact of simulation-based training for laparoscopic surgery but focused their evaluation on process measures such as knowledge, skill time, skill process etc. with only one study on patient effect.[37] Process measures are fundamental to performance of the procedure, i.e., how long it took, but gives no indication of the quality of procedure performance. The review reported here focuses on prospective, randomized and blinded clinical studies specifically on PBP simulation training. The aim of this review is to evaluate the impact of this approach to learning clinical skills in comparison to the traditional approach to training.

2. Materials and Methods

Study identification and evaluation

A systematic review of the literature was conducted using the PubMed, Cochrane library's Central, EMBASE, MEDLINE and Scopus databases (Supplementary Material appendix XX). We searched from inception of the databases up to 1st March 2020. All the references of key reviews on training were also screened. Key words used for the research were: "Proficiency-based AND progression AND training, Proficiency AND based AND progression, Proficiency-based AND training". Thise systematic review is~~was~~ reported in accordance with~~ing to~~ the Preferred r~~R~~eporting i~~I~~tems for s~~S~~ystematic reviews and m~~M~~eta-analyses protocols (PRISMA-P) guidelines[38]. The current study is~~and~~ registered within~~in~~ the international prospective register of systematic reviews (PROSPERO, ID XXX).

Initial screening, eligibility criteria and risk of bias assessment

After identifying all eligible studies ~~eligible~~, 2 independent reviewers (MA, ST) screened all titles and abstracts (or full text, for further clarification~~ if doubt~~) for inclusion in the study. Literature reviews, editorial, comments, and non PBP-based studies (other than as a control condition) ~~studies, non-comparative studies, non-prospective studies~~ were excluded at the initial screening (Figure 1). O~~After eligibility evaluation, o~~nly ~~those~~ studies that used an~~using~~ objective performance based metrics~~ based~~ and a PBP methodology were included for the final quantitative synthesis.[15,24,30–32,34,39–42] Any disagreements about eligibility were resolved by discussion between the two investigators until consensus was reached.

Methodological quality of the included studies was graded using the Medical Education Research Study Quality Instrument (MERSQI).[43] Three investigators (EM, SP and AGG) independently assessed the risk of bias for all studies and the inter-rater reliability (IRR) of the assessors was calculated (i.e., IRR = Agreements/Agreements + Disagreements).[7]

Intervention and comparison arms

The training tasks/procedures considered for the meta-analytic comparison were categorized as follows: medical procedure, surgical procedure, basic skill and clinical communication skill. Intraoperative patient performance was considered as the direct or post-training impact on patients of a training pathway.

For meta-analytic evaluation, the PBP simulation-based training arm was considered as the experimental arm. The group which received a non-PBP simulation-based training represented the comparison arm. For both arms, studies including any simulators, as well as those including virtual reality simulators (or other technology-enhanced training models), box model or human cadaveric models, were considered eligible.

Outcomes definition

PBP training has been previously described in detail.[27,28,44] According to PBP-related definitions, metrics are explicitly defined units of measurement that characterize elements of procedure/task performance and are scored in a binary fashion (i.e., occurred/did not occur). The metrics are quantitative assessments and are used for objective evaluations to make comparisons or to track performance. This included~~We considered~~ both performance errors and steps as metrics since~~as~~ they were explicitly defined as occurring-not occurring, and were objectively assessable. Error was defined as a 'deviation from the optimal performance'. Steps were defined as component tasks~~,~~ and the series was aggregated since they ~~of which~~ constitute~~s~~ the completion of a specific procedure.[31] Only the studies that specified those parameters in their analysis were included within~~for~~ the qualitative analysis. This included~~ as were~~ studies that used ~~the~~ metrics to define a proficiency benchmark by which~~that~~ trainees were required to demonstrate the benchmark before training was deemed completed. Time was also considered as additional outcome. Assessment by Likert scales

~~assessment~~ were not included in the current analyses, because of the potential for inherent ambiguity.~~since they are not based on objective performance evaluation.~~

All the study outcomes ~~have been categorized as~~ meeting these criteria are shown in Table 2. The primary outcome used for pooled meta-analysis was the number of procedural errors performed (based on the objective deconstruction of the respective task/procedure), since errors provide an objective measure of performance quality.[21,27,28,31] Secondary outcomes were the number of steps performed and time to completion~~e~~ of the task/procedure, both of which provide measures of process for ~~of~~ task/procedure performance.

Data synthesis and statistical analysis

Data not suitable for meta-analytic evaluation was presented in narrative fashion (qualitative analysis). Reported results for continuous outcomes were pooled using biased corrected standardized mean difference (SMD) (Hedges' g effect size) according to previous established methodology.[37,45] Thus, the bias corrected SMD and odds ratio (OR) were used to compare continuous and dichotomous variables, respectively. Additionally, for continuous outcomes, ratio of means (ROM) was applied to provide an estimation of the pooled effect of PBP on the considered outcomes.[46,47] All results were reported with 95% confidence intervals. Pre-planned subgroups analyses were performed in studies with or without intraoperative patient performance assessment.

Heterogeneity between studies was measured using the I^2 statistic [48] and the between-study variance (t^2) from the random-effect analyses. I^2 values >50% indicate large inconsistency. In case of large heterogeneity, random effect models (using the DerSimonian and Laird approach[49]) were prioritized. For the assessment of publication bias and small study effects, values of the SMD or OR were plotted against their standard error in a contour-enhanced funnel plot. Furthermore, Eggers asymmetry test[50] was used to explore statistically ~~explore~~ the presence of publication bias. Statistical significance for all analysis was defined as two-sided $p < 0.05$. Statistical analysis was performed with the R software (version 3.6.3; http://www.r-project.org/). Unless otherwise indicated all models have allowed for different effect sizes (random-effects).

3. Results

3.1 Study selection flow-chart

Figure 1 shows the flow of studies through the screening process. First, 519 papers were blindly screened by two reviewers (MA, ST) by reading all the titles and abstracts. After the first screening, 463 records were included for further evaluation based on pre-defined eligibility criteria. Of these, 40 studies were considered eligible for final inclusion in qualitative analysis. Here, final evaluation for the inclusion in the quantitative synthesis was carried out by 3 reviewers (AG, EM, SP). At the end of the process, 12 manuscripts have been included for the meta-analysis.

3.2 Study quality and Risk of bias

The Supplementary Material appendix XX summarises the quality criteria assessed for each RCT using the MERSQI tool. The overall methodological quality of the studies was high, with all the studies having low risk of bias. Notably, the overall mean score of the RCTs was 15.5 (range 14.5 and 17). The mean IRR of quality scores between assessors was 94.6% (range 80-100%).

3.3 Evidence synthesis

Tables 1 and 2 summarize general and design characteristics of the selected studies. Primary analysis included 12 papers for qualitative review and quantitative synthesis. The final screened manuscripts reported outcomes based on 5 full surgical procedures, 3 surgical skill tasks (i.e., steps or part of a procedure, knotting and/or suturing), 3 non-surgical medical procedures and 1 clinical communication skill task~~s~~. Overall, 12 attendings in practice (1 study), 161 residents (10 studies) and 66 medical students (2 studies) were evaluated in the included RCTs. Of these, 85 ~~and 76~~ participants had been allocated to a~~undergoing respectively~~ PBP condition and n=76 were in an non-PBP-based training pathways.~~ received final procedure performance assessment.~~ According to the primary outcome (i.e. number of errors), 9 studies (199 participants) were included in the quantitative synthesis (i.e. meta-analysis). For steps, time and proficiency assessment on the procedure, 6 (134 participants), 6 (100 participants) and 3 (110 events) studies were respectively included in the quantitative comparisons. [I DON'T UNDERSTAND THE ABOVE SENTENCE.]

In quantitative synthesis testing for procedural errors, a pooled meta-analysis on 199 trainees was conducted (Fig. 2a-b), using random-effects models. Overall, PBP training reduced the number of errors when compared to standard training (SMD -2.68, 95% CI: -3.52; -1.83; p < 0.001 ~~at random effects model~~). In a~~At~~ ROM analysis, PBP was estimated to reduce the mean rate of errors by~~of~~ approximately 60%, when compared to standard training (ROM 0.42, 95% CI: 0.32; 0.55; p < 0.001 ~~at random effects model~~). Funnel plot and Eggers' linear regression estimates both showed evidence for ~~presence of~~ potential publications bias (Supplementary material appendix XX). In subgroup analyses, focusing on studies with intraoperative patient performance assessment (n = 87), PBP training outperformed standard training (SMD -3.11, 95% CI: -4.54; -1.68; p < 0.001 ~~at random effects model~~), with an estimated reduction in mean rates of errors of 62% (ROM 0.38, 95% CI: 0.25; 0.58; p < 0.001 ~~at random effects model~~).

For~~Within~~ secondary outcomes, in quantitative synthesis testing for number of steps completed, a pooled meta-analysis on 134 trainees was conducted. Overall, trainees who completed PBP training performed more procedural steps than those who completed a standard training pathway (SMD 3.46, 95% CI: 2.13; 4.79; p < 0.001 ~~at random effects model~~) (Fig. 3a). At ROM analysis, PBP increased the mean rate of steps performed by an average of 43%, when compared to standard training (ROM 1.46, 95% CI: 1.21; 1.77; p <0 .001 ~~at random effects model~~) (Fig. 3b). Funnel plot and Eggers' linear regression estimates recorded a marginal effect for~~of~~ potential publications bias (Supplementary material appendix XX). In the two studies reporting the effect of PBP on steps performed in intraoperative patient procedure, PBP was shown~~confirmed~~ to increase the number of steps performed (SMD 3.90, 95% CI: 1.79; 6.02; p < 0.001 ~~at random effect~~) but in ROM analysis such a difference failed to achieve statistical significance (ROM 1.28, 95% CI: 0.94; 1.74; p = 0.1 ~~at random effect~~).

In quantitative synthesis testing for procedural time, a pooled meta-analysis on 100 trainees was conducted. Overall, trainees who completed PBP training performed more procedural steps than those who completed a standard training pathway (SMD -0.93, 95% CI: -1.55; -0.30; p = 0.003~~ at random effect model~~) (Fig. 3c). As expected, reduction of procedural time was less pronounced compared to other outcomes, such as the number of errors or steps completed. Indeed, at ROM analysis, PBP reduced the mean procedural time by~~of~~ approximately 15%, when compared to standard training (ROM 0.85, 95% CI: 0.75-0.96, p = 0.009~~ at random effects model~~) (Fig 3d). Funnel plot and Eggers' linear regression estimates demonstrate an~~s ~~absence of potential publications bias (Supplementary material appendix XX). In subgroup analyses focusing on studies with intraoperative patient procedure assessment, PBP training slightly outperformed standard training (SMD -0.86, 95% CI: -1.65, -0.08; p=0.03~~ at random effects model~~), with an estimated decrease in mean completion time of 19% (ROM 0.81, 95% CI 0.65; 1.01; p = 0.06~~ at random effects model~~).

Finally in the~~Lastly, at~~ quantitative synthesis testing for the rate of proficiency benchmark achievement on the procedure, a pooled meta-analysis on 110 trainees was conducted (Supplementary material appendix XX). Overall, trainees who completed PBP were more likely to reach the proficiency benchmark when compared to those who completed a standard training pathway (OR 8.67, 95% CI: 2.52; 29.77; p < 0.001 using a~~at~~ fixed effects model). Funnel plot and Eggers' linear regression estimates demonstrated an absence of potential publications bias (Supplementary material appendix XX). Only one study reported results based on intraoperative patient procedure assessment, and it confirmed~~ing~~ the protective effect of PBP training on achieving the final proficiency benchmark (OR 7.50, 95% CI 1.31; 43.03; p = 0.02 ~~a~~in ~~at~~ fixed effects model).

4. Discussion

Surgery and procedure-based medical treatments require standardized, and precise training that can guarantee high-quality healthcare~~ standard~~s[51,52]. In this context, during the last two decades, several studies have underlined the importance of implementing validated training curricula[53,54], based on technology enhanced simulation training, in order to achieve a high standard of skills before starting with clinical practice on patients~~real cases~~. However, a technology enhanced simulation training pathway may not be sufficient if it is not supported by an effective teaching methodology, that is~~should be~~ based on the fundamental concepts of deliberate practice[29] and PBP.[27]

PBP simulation training was first reported ~~as being used~~ in a clinical study in 2002 and the methodology was described in detail in 2005[44] with subsequent reports.[27,28] In this systematic review of peer-reviewed, published, prospective, randomised and blinded clinical studies we report the meta-analysis and results from 12 studies.~~ that meet these criteria with extractable data summaries.~~ The studies were carried out in the following medical specialities: laparoscopic surgery, arthroscopic surgery, interventional cardiology/endovascular interventions, anaesthesia and clinical communication skills. As measured with the MERSQI instrument the quality of the studies was high. ~~Furthermore, findings from all of the studies were homogeneous and with intervention effects also all in the same direction.~~

PBP training consistently showed significant improvements in performance by trainees. Significant improvements in performance/procedure time and procedure steps completed were observed. The largest and most consistent improvements however were found for error performance, particularly intra-operative errors on ~~real~~ patients. In studies that evaluated~~included or objectively measured~~ intraoperative errors, we found a 62% reduction in comparison to the standard training group. For studies that assessed performance outside the operating room, or clinical environment, we found a 50% reduction in objectively assessed performance errors.

The number of steps completed by the clinician is fundamental to the safe completion of the procedure,~~.~~ and~~Similarly,~~ the completion of the procedure will inevitably take a certain amount of time. These two measures, however, provide~~give very~~ little substantiation regarding~~of~~ the quality of performance. For example, all of the steps of a procedure may be completed, but done badly. Likewise, a procedure can be performed quickly but unsafely, or phases of the procedure can be omitted resulting in faster completion times.[27,28,31] Neither measures give a reliable indication of the quality of the operator's performance. In contrast, objectively assessed performance error in the PBP methodology give~~s~~ direct, objective, transparent and fair measures of quality.

T~~Of note, t~~he impact of PBP training was greatest on objectively assessed intraoperative performance errors. There was however~~ only~~ one study which directly assessed the impact of PBP training on a clinical outcome. Srinivasan et al.[33]~~,~~ assessed the impact of PBP simulation training on the effectiveness and success of epidural analgesia administration during labour. They found that the PBP trained group had a 54% lower epidural failure rate than the simulation trained group.

Traditionally, it has been (incorrectly) believed that only about ~10% of what was learned in the training environment transferred to real-world performance.[55] However, a number of studies have shown substantial transfers of training skills. For example, in the training of laparoscopic surgical skills ~~A transfer of training study of laparoscopic surgical skills refuted this claim and demonstrated that there was~~ a 26% improvement in performance process measures (e.g. performance time) and a 42% improvement in performance quality measures (i.e., performance errors) have been reported.[56] Others have ~~It was however~~ believed that some of these observed effects of transfer from training would inevitably be lost, though~~.~~ t~~T~~he results from the Srinivasan et al.[33]~~,~~ study suggest that this is unlikely to be the case~~they may not be lost~~. Further~~more~~, research has also indicated~~now shows~~ that ~~the~~ objectively assessed skills of clinicians significantly affect the~~impact on~~ procedure outcome.[57,58] In this context, Birkmeyer et al.[57], objectively assessed the technical skills of experienced bariatric surgeons. On the basis of this assessment, surgeons were stratified into four quartiles, i.e., surgeons who were performing best, two middle quartiles and surgeons who were performing least well~~worst~~. All of the bariatric patients that these surgeons operated on over the next six years were studied. They found the best performing surgeons~~,~~ had a surgical complication rate of 4.2% in comparison to the bottom quartile group which had an 11.4% complication rate: a difference of 63%; an infection rate of 1.04% in comparison to 4.6%: a difference of 77%. Although the overall mortality rate for the study was low, the best performing surgeons had a rate of 0.05% in comparison to 0.26% in the bottom quartile group: a difference of 81%. These data strongly corroborated the correlation between skills acquisition and intraoperative performance and patient-related outcomes.

The effectiveness of the PBP simulation training is probably accounted for by a number of factors. The first is that the performance characteristics on which training is based~~to be trained~~ are derived from very experienced and practising clinicians. They identify the characteristics and performances necessary for ~~a~~ trainees at the start of their learning curve, and hence provide ~~to successfully complete~~ a reference approach to the successful performance of the procedure [17,19–21] and provide the basis for a~~these~~ performance metrics that can be~~are then~~ validated.[39,44]

Once ~~satisfactorily~~ validated, a proficiency benchmark is established based on the mean performance of experienced practitioners.[5,15,27,30–32,34,44] Another fundamental aspect of PBP training is that the detailed metrics are used to provide~~give~~ the trainees with objective, transparent and constructive feedback on their performance, thus~~,~~ affording trainees the opportunity to engage in~~ a~~ deliberate practice training rather than repeated practice.[29]

Lastly, PBP training is not complete until the trainee has demonstrated - a level of proficiency based on pre~~the quantitatively~~ defined~~ proficiency~~ benchmarks. The benchmark is based on the mean of the objectively assessed performance of very experienced and proficient clinicians performing the exact same task/procedure. Thus, the trainee knows what to do, with what instruments and in which order. They also have demonstrated that they can adequately undertake the task under conditions of a~~ado it on the~~ simulation or training model, and that they can achieve to a~~the~~ quantitatively defined proficiency benchmark. The pre-trained novice ~~has however~~ never completes~~d~~ the medical procedure on a live ~~real~~ patient until they have shown that they can adequately perform the task within a training context.~~ their first supervised case~~. Evidence reviewed here suggests that PBP ensure that trainees~~they~~ are significantly better prepared than more traditionally trained clinicians.

Conclusion

Our systematic review and meta-analysis of RCTs confirms that PBP training improves trainees' performances when compared to standard simulation-based training. Notably, PBP decreases procedural errors by 60% compared to conventional/traditional training and such a positive impact on trainees' performances is higher when focusing on intraoperative performance assessment. These results reinforce the need to fully implement PBP methodology in surgical and procedure-based medical treatments training pathways.

REFERENCES

1. Satava RM. Virtual reality surgical simulator. *Surg Endosc*. 1993;7(3):203-205. doi:10.1007/BF00594110
2. Salas E, Bowers CA, Rhodenizer L. It is not how much you have but how you use it: toward a rational use of simulation to support aviation training. *Int J Aviat Psychol*. 1998;8(3):197-208. doi:10.1207/s15327108ijap0803_2
3. Webster CS. The nuclear power industry as an alternative analogy for safety in anaesthesia and a novel approach for the conceptualisation of safety goals. *Anaesthesia*. 2005;60(11):1115-1122. doi:10.1111/j.1365-2044.2005.04301.x
4. Gaba DM, DeAnda A. A comprehensive anesthesia simulation environment: re-creating the operating room for research and training. *Anesthesiology*. 1988;69(3):387-394.
5. Seymour NE, Gallagher AG, Roman SA, et al. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Ann Surg*. 2002;236(4):454-458. doi:10.1097/00000658-200210000-00008
6. Seymour NE, Gallagher AG, Roman SA, O’Brien MK, Andersen DK, Satava RM. Analysis of errors in laparoscopic surgical procedures. *Surg Endosc*. 2004;18(4):592-595. doi:10.1007/s00464-002-8927-2
7. Kazdin AE. *Behavior Modification in Applied Settings, 5th Ed.* Belmont, CA, US: Thomson Brooks/Cole Publishing Co; 1994.
8. Gallagher AG, O’Sullivan GC, Leonard G, Bunting BP, McGlade KJ. Objective structured assessment of technical skills and checklist scales reliability compared for high stakes assessments. *ANZ J Surg*. 2014;84(7-8):568-573. doi:10.1111/j.1445-2197.2012.06236.x
9. Gallagher AG, Ryu RKN, Pedowitz RA, Henn P, Angelo RL. Inter-rater Reliability for Metrics Scored in a Binary Fashion-Performance Assessment for an Arthroscopic Bankart Repair. *Arthroscopy*. 2018;34(7):2191-2198. doi:10.1016/j.arthro.2018.02.007
10. Gallagher AG, Satava RM. Virtual reality as a metric for the assessment of laparoscopic

psychomotor skills. Learning curves and reliability measures. *Surg Endosc*. 2002;16(12):1746-1752. doi:10.1007/s00464-001-8215-6

11. Gallagher AG, Richie K, McClure N, McGuigan J. Objective psychomotor skills assessment of experienced, junior, and novice laparoscopists with virtual reality. *World J Surg*. 2001;25(11):1478-1483. doi:10.1007/s00268-001-0133-1

12. Jordan JA, Gallagher AG, McGuigan J, McClure N. Virtual reality training leads to faster adaptation to the novel psychomotor restrictions encountered by laparoscopic surgeons. *Surg Endosc*. 2001;15(10):1080-1084. doi:10.1007/s004640000374

13. Gallagher, A. E., Hughes, C., Reinhardt-Rutland, A., McGuigan, J., & McClure N. Transfer of skill from'virtual reality' (VR): A case-control comparison of traditional and VR training in laparoscopic skill acquisition. *Minim Invasive Ther Allied Technol*. 2000.

14. Gallagher AG, McClure N, McGuigan J, Crothers I, Browning J. Virtual reality training in laparoscopic surgery: a preliminary assessment of minimally invasive surgical trainer virtual reality (MIST VR). *Endoscopy*. 1999;31(4):310-313. doi:10.1055/s-1999-15

15. Ahlberg G, Enochsson L, Gallagher AG, et al. Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. *Am J Surg*. 2007;193(6):797-804. doi:10.1016/j.amjsurg.2006.06.050

16. Mascheroni J, Mont L, Stockburger M, Patwala A, Retzlaff H, Gallagher AG. International expert consensus on a scientific approach to training novice cardiac resynchronization therapy implanters using performance quality metrics. *Int J Cardiol*. 2019;289:63-69. doi:10.1016/j.ijcard.2019.04.036

17. Angelo RL, Ryu RKN, Pedowitz RA, Gallagher AG. Metric Development for an Arthroscopic Bankart Procedure: Assessment of Face and Content Validity. *Arthroscopy*. 2015;31(8):1430-1440. doi:10.1016/j.arthro.2015.04.093

18. Angelo RL, Tauro J, St Pierre P, et al. Arthroscopic Rotator Cuff Repair Metrics: Establishing Face, Content, and Construct Validity in a Cadaveric Model. *Arthrosc J*

Arthrosc Relat Surg Off Publ Arthrosc Assoc North Am Int Arthrosc Assoc. 2020;36(1):71-79.e1. doi:10.1016/j.arthro.2019.07.016

19. Kojima K, Graves M, Taha W, Cunningham M, Joeris A, Gallagher AG. AO international consensus panel for metrics on a closed reduction and fixation of a 31A2 pertrochanteric fracture. *Injury*. 2018;49(12):2227-2233. doi:10.1016/j.injury.2018.09.019
20. Mascheroni J, Mont L, Stockburger M, Patwala A, Retzlaff H, Gallagher AG. A validation study of intraoperative performance metrics for training novice cardiac resynchronization therapy implanters. *Int J Cardiol*. 2020;307:48-54. doi:10.1016/j.ijcard.2020.02.003
21. Crossley R, Liebig T, Holtmannspoetter M, et al. Validation studies of virtual reality simulation performance metrics for mechanical thrombectomy in ischemic stroke. *J Neurointerv Surg*. 2019;11(8):775-780. doi:10.1136/neurintsurg-2018-014510
22. Angelo RL, Pedowitz RA, Ryu RKN, Gallagher AG. The Bankart Performance Metrics Combined With a Shoulder Model Simulator Create a Precise and Accurate Training Tool for Measuring Surgeon Skill. *Arthrosc J Arthrosc Relat Surg Off Publ Arthrosc Assoc North Am Int Arthrosc Assoc*. 2015;31(9):1639-1654. doi:10.1016/j.arthro.2015.04.092
23. Angelo RL, Ryu RKN, Pedowitz RA, Gallagher AG. The Bankart Performance Metrics Combined With a Cadaveric Shoulder Create a Precise and Accurate Assessment Tool for Measuring Surgeon Skill. *Arthrosc J Arthrosc Relat Surg Off Publ Arthrosc Assoc North Am Int Arthrosc Assoc*. 2015;31(9):1655-1670. doi:10.1016/j.arthro.2015.05.006
24. Pedowitz RA, Nicandri GT, Angelo RL, Ryu RKN, Gallagher AG. Objective Assessment of Knot-Tying Proficiency with the Fundamentals of Arthroscopic Surgery Training Program Workstation and Knot Tester. *Arthrosc - J Arthrosc Relat Surg*. 2015;31(10):1872-1879. doi:10.1016/j.arthro.2015.06.021
25. Ritter EM, McClusky DA 3rd, Gallagher AG, Smith CD. Real-time objective assessment of knot quality with a portable tensiometer is superior to execution time for assessment of laparoscopic knot-tying performance. *Surg Innov*. 2005;12(3):233-237.

doi:10.1177/155335060501200308

26. Angelo RL, Pedowitz RA, Ryu RKN, Gallagher AG. The Bankart Performance Metrics Combined With a Shoulder Model Simulator Create a Precise and Accurate Training Tool for Measuring Surgeon Skill. *Arthroscopy*. 2015;31(9):1639-1654. doi:10.1016/j.arthro.2015.04.092

27. Gallagher AG, O’Sullivan GC. *Fundamentals of Surgical Simulation: Principles and Practice*. Springer Publishing Company, Incorporated; 2011. doi:10.1007/978-0-85729-763-1

28. Gallagher AG. Metric-based simulation training to proficiency in medical education:- what it is and how to do it. *Ulster Med J*. 2012;81(3):107-113.

29. Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad Med*. 2004;79(10 SUPPL.):70-81. doi:10.1097/00001888-200410001-00022

30. Van Sickle KR, Ritter EM, Baghai M, et al. Prospective, randomized, double-blind trial of curriculum-based training for intracorporeal suturing and knot tying. *J Am Coll Surg*. 2008;207(4):560-568. doi:10.1016/j.jamcollsurg.2008.05.007

31. Angelo RL, Ryu RKN, Pedowitz RA, et al. A Proficiency-Based Progression Training Curriculum Coupled With a Model Simulator Results in the Acquisition of a Superior Arthroscopic Bankart Skill Set. *Arthroscopy*. 2015;31(10):1854-1871. doi:10.1016/j.arthro.2015.07.001

32. Cates CU, Lönn L, Gallagher AG. Prospective, randomised and blinded comparison of proficiency-based progression full-physics virtual reality simulator training versus invasive vascular experience for learning carotid artery angiography by very experienced operators. *BMJ Simul Technol Enhanc Learn*. 2016;2(1):1-5. doi:10.1136/bmjstel-2015-000090

33. Srinivasan KK, Gallagher A, O’Brien N, et al. Proficiency-based progression training: an “end to end” model for decreasing error applied to achievement of effective epidural

analgesia during labour: a randomised control study. *BMJ Open*. 2018;8(10):e020099. doi:10.1136/bmjopen-2017-020099

34. Breen D, O’Brien S, McCarthy N, Gallagher A, Walshe N. Effect of a proficiency-based progression simulation programme on clinical communication for the deteriorating patient: a randomised controlled trial. *BMJ Open*. 2019;9(7):e025992. doi:10.1136/bmjopen-2018-025992

35. Sutherland LM, Middleton PF, Anthony A, et al. Surgical simulation: a systematic review. *Ann Surg*. 2006;243(3):291-300. doi:10.1097/01.sla.0000200839.93965.26

36. Gurusamy K, Aggarwal R, Palanivelu L, Davidson BR. Systematic review of randomized controlled trials on the effectiveness of virtual reality training for laparoscopic surgery. *Br J Surg*. 2008;95(9):1088-1097. doi:10.1002/bjs.6344

37. Zendejas B, Brydges R, Hamstra SJ, Cook DA. State of the evidence on simulation-based training for laparoscopic surgery: a systematic review. *Ann Surg*. 2013;257(4):586-593. doi:10.1097/SLA.0b013e318288c40b

38. PRISMA statement. http://www.prisma-statement.org.

39. Ahmed OM, O’Donnell BD, Gallagher AG, Shorten GD. Development of performance and error metrics for ultrasound-guided axillary brachial plexus block. *Adv Med Educ Pract*. 2017;8:257-263. doi:10.2147/AMEP.S128963

40. Palter VN, Grantcharov TP. Development and validation of a comprehensive curriculum to teach an advanced minimally invasive procedure: a randomized controlled trial. *Ann Surg*. 2012;256(1):25-32. doi:10.1097/SLA.0b013e318258f5aa

41. Peeters SHP, Akkermans J, Slaghekke F, et al. Simulator training in fetoscopic laser surgery for twin-twin transfusion syndrome: a pilot randomized controlled trial. *Ultrasound Obstet Gynecol Off J Int Soc Ultrasound Obstet Gynecol*. 2015;46(3):319-326. doi:10.1002/uog.14916

42. Jensen UJ, Jensen J, Ahlberg G, Tornvall P. Virtual reality training in coronary angiography

and its transfer effect to real-life catheterisation lab. *EuroIntervention J Eur Collab with Work Gr Interv Cardiol Eur Soc Cardiol*. 2016;11(13):1503-1510. doi:10.4244/EIJY15M06_05

43. Reed DA, Cook DA, Beckman TJ, Levine RB, Kern DE, Wright SM. Association between funding and quality of published medical education research. *JAMA*. 2007;298(9):1002-1009. doi:10.1001/jama.298.9.1002

44. Gallagher AG, Ritter EM, Champion H, et al. Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. *Ann Surg*. 2005;241(2):364-372. doi:10.1097/01.sla.0000151982.85062.80

45. Cook DA, Hatala R, Brydges R, et al. Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. *JAMA*. 2011;306(9):978-988. doi:10.1001/jama.2011.1234

46. Friedrich JO, Adhikari NKJ, Beyene J. Ratio of means for analyzing continuous outcomes in meta-analysis performed as well as mean difference methods. *J Clin Epidemiol*. 2011;64(5):556-564. doi:10.1016/j.jclinepi.2010.09.016

47. Friedrich JO, Adhikari NKJ, Beyene J. The ratio of means method as an alternative to mean differences for analyzing continuous outcome variables in meta-analysis: a simulation study. *BMC Med Res Methodol*. 2008;8:32. doi:10.1186/1471-2288-8-32

48. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327(7414):557-560. doi:10.1136/bmj.327.7414.557

49. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177-188. doi:10.1016/0197-2456(86)90046-2

50. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997;315(7109):629-634. doi:10.1136/bmj.315.7109.629

51. Palagonia E, Mazzone E, De Naeyer G, et al. The safety of urologic robotic surgery depends on the skills of the surgeon. *World J Urol*. August 2019. doi:10.1007/s00345-019-02901-9

52. Puliatti S, Mazzone E, Dell’Oglio P. Training in robot-assisted surgery. *Curr Opin Urol*. 2020;30(1):65-72. doi:10.1097/MOU.0000000000000687

53. Brook NR, Dell’Oglio P, Barod R, Collins J, Mottrie A. Comprehensive training in robotic surgery. *Curr Opin Urol*. November 2018:1. doi:10.1097/MOU.0000000000000566

54. Mazzone E, Dell’Oglio P, Mottrie A. Outcome report of the first ERUS robotic urology curriculum-trained surgeon in Turkey: the importance of structured and validated training programs for global outcome improvement. *Turk J Urol*. 2019;(DOI: 10.5152/tud.2019.19019). doi:10.5152/tud.2019.19019

55. Holton III E, Baldwin T. *Improving Learning Transfer in Organizations*. John Wiley & Sons; 2003.

56. Gallagher AG, Seymour NE, Jordan-Black J-A, Bunting BP, McGlade K, Satava RM. Prospective, randomized assessment of transfer of training (ToT) and transfer effectiveness ratio (TER) of virtual reality simulation training for laparoscopic skill acquisition. *Ann Surg*. 2013;257(6):1025-1031. doi:10.1097/SLA.0b013e318284f658

57. Birkmeyer JD, Finks JF, O’Reilly A, et al. Surgical skill and complication rates after bariatric surgery. *N Engl J Med*. 2013;369(15):1434-1442. doi:10.1056/NEJMsa1300625

58. Curtis NJ, Foster JD, Miskovic D, et al. Association of Surgical Skill Assessment With Clinical Outcomes in Cancer Surgery. 2020:1-10. doi:10.1001/jamasurg.2020.1004

Table 1. General characteristics of 12 randomized clinical trials studies included in the final qualitative analysis of the systematic review.

Study	Design	Subjects N; Type	Comparison arm*	Task/Procedure Trained	Intraoperative patient performance	Outcomes Compared	Other scale used	MERS QI
Ahlberg et al	RCT	13; Residents	Vr ST	Laparoscopic Cholecystectomies	Yes	E	--	16
Ahmed et al.	RCT	18; Medicine Students	Self-guided practice	Ultrasound-Guided Peripheral Nerve Block	No	S, E	--	15
Angelo et al.	RCT	44; Residents	Simulation-based ST	Arthroscopic Bankart Procedure	Yes	S, E, T	--	16
Breen et al.	RCT	90; Medicine and nursing students	ST	Clinical Communication	No	S, E	--	15
Cates et al.	RCT	12; Attendings	Vr ST	Carotid Artery Angiography	Yes	T, E	--	15
Jensen et al.	RCT	16; Residents	Vr ST	Coronary Angiography	No	T, E, S	--	17
Palter, et al.	RCT	25; Residents	Vr and cadaver ST	Laparoscopic Right Colectomy	Yes	S	OSATS	16
Pedowitz et al.	RCT	44; Residents	Simulation-based ST	Knot-Tying	No	E	--	14.5
Peeters, et al.	RCT	10; Residents	Simulation-based ST	Fetoscopy Laser Surgery	No	S, T	--	16.5
Seymour et al.	RCT	16; Residents	Vr ST	Laparoscopic Cholecystectomy	Yes	E, T	--	15
Srinivasan et al.	RCT	17; Residents	Simulation-based ST	Epidural Analgesia	Yes	E	GRS, TSCL	17
Van Sickle et al.	RCT	22; Residents	Simulation-based ST	Intracorporeal Suturing and Knot	Yes	T, E	--	14.5

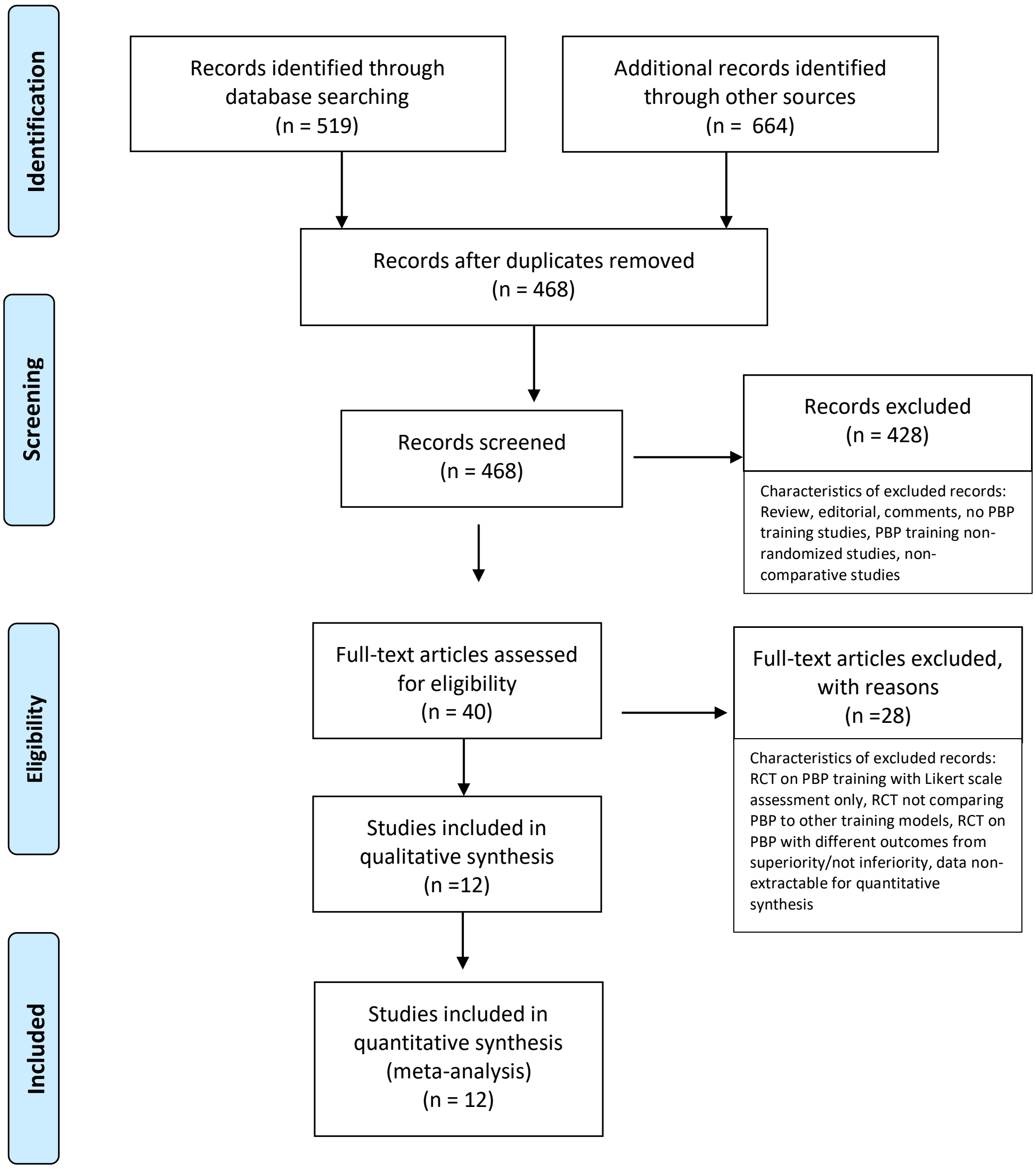
* If more than one arm was included in the original study, the one with standard or classic training or no additional features was considered as comparison arm
E: errors; T: time; S: steps; ST: standard training; Vr: Virtual reality ; OSATS: objective structured assessment of technical skills; GEARS: global evaluative assessment of robotic skills; GRS: Global rating scales; TSCL: Task-specific checklists; RCT: randomized controlled trial.

Table 2. Baseline characteristics of included studies according to participants data and outcomes measures.

		All Outcomes	All outcomes - PBP group	All outcomes - Control group	Time	Errors	Steps	Proficiency
Feature	**Subgroup**	**No. Studies (No. Participants)**	**No. Studies (No. Participants)**	**No. Studies (No. Participants)**	**No. Studies (No. Participants)**	**No. Studies (No. Participants)**	**No. Studies (No. Participants)**	**No. Studies (No. Participants)**
All		12(239)	12(127)	12(112)	6(100)	10(211)	6(134)	3(102)
Design	**RCT**	12	12	12	6	10	6	3
Participants	**Medical/nursing Students**	2(66)	2(36)	2(30)	0	2(66)	2(66)	1(48)
	Residents	9(161)	9(85)	9(76)	5(88)	7(133)	4(68)	2(54)
	Physicians in practice	1(12)	1(6)	1(6)	1(12)	1(12)	0	0
Task or Procedure	**Skill**	3(70)	3(37)	3(33)	1(22)	3(70)	1(18)	1(30)
	Surgical procedure	4(63)	4(31)	4(32)	3(50)	3(40)	2(34)	0
	Medical procedure	4(58)	4(44)	4(14)	2(28)	3(53)	2(34)	1(24)
	Not medical procedure	1(48)	1(25)	1(23)	0	1(48)	1(48)	1(48)
Clinical Relevance	**Present**	7(117)	7(73)	7(44)	4(84)	6(99)	2(42)	1(24)
	Absent	5(122)	5(54)	5(68)	2(16)	4(112)	4(92)	2(78)
Outcomes	**Satisfaction, aptitude etc**	0	0	0	0	0	0	0
	Knowledge, skills	3(70)	3(37)	3(33)	1(22)	3(70)	1(18)	0
	Behavior	8(157)	8(79)	8(78)	5(78)	6(129)	5(116)	3(102)
	Patient/health system outcomes	1(12)	1(11)	1(1)	0	1(12)	0	0

RCT: randomized clinical trial

Figure 1. Flow-chart of studies through the screening process according to the PRISMA methodology



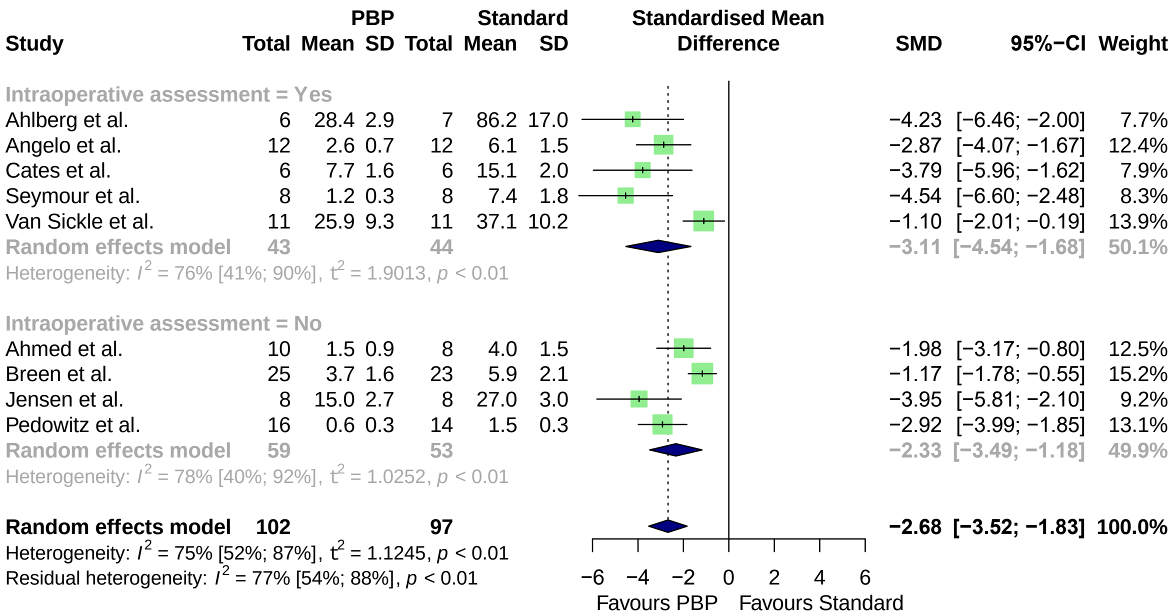


Figure 2A. Standardized mean difference between studies assessing the effect of PBP vs standard training on procedural errors.

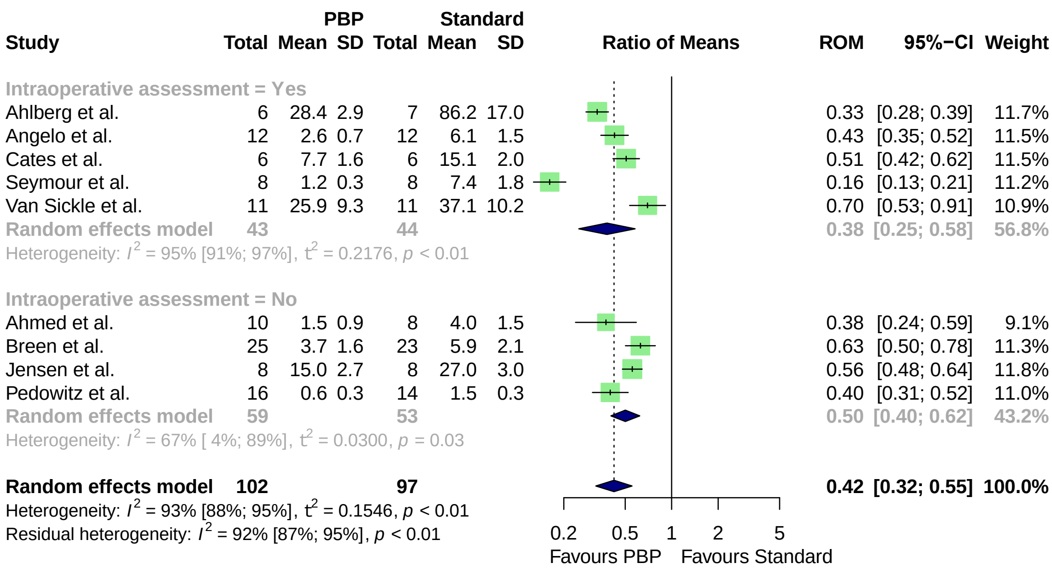
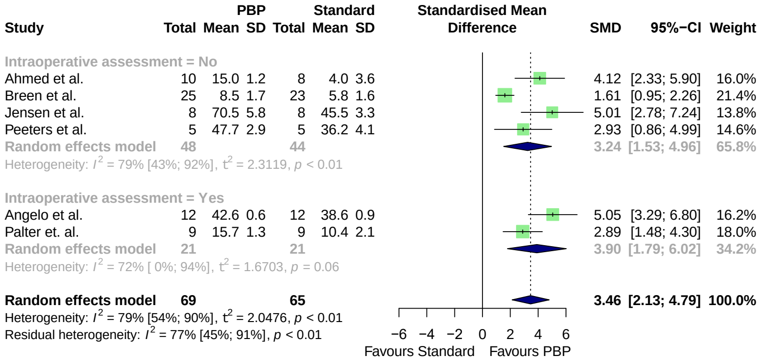


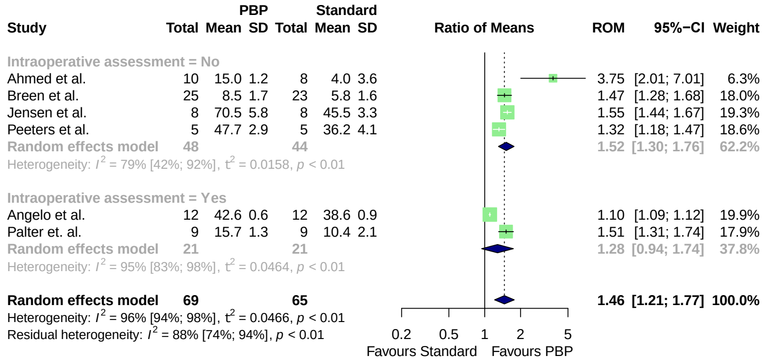
Figure 2B. Ratio of means between studies assessing the effect of PBP vs standard training on procedural errors.

Figure 3. Standardized mean difference and ratio of means between studies assessing the effect of PBP vs standard training on procedural steps (A-B) and procedural time (C-D).

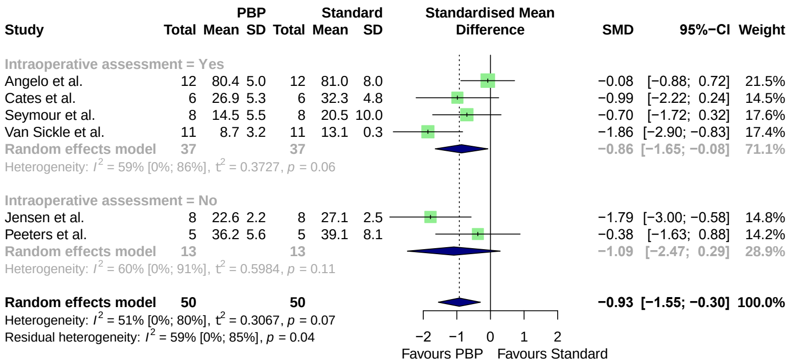
A



B



C



D